

## A SYSTEMATIC COMPARISON OF DIRECT AND IMAGE-BASED GEOREFERENCING IN CHALLENGING URBAN AREAS

S. Cavegn<sup>a, b</sup>, S. Nebiker<sup>a</sup>, N. Haala<sup>b</sup>

<sup>a</sup> Institute of Geomatics Engineering, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland - (stefan.cavegn, stephan.nebiker)@fhnw.ch

<sup>b</sup> Institute for Photogrammetry, University of Stuttgart, Germany - norbert.haala@ifp.uni-stuttgart.de

Commission I, ICWG I/Va

**KEY WORDS:** Image, Sequence, Stereo, Mobile Mapping, Direct, Integrated, Georeferencing, Sensor Orientation

### ABSTRACT:

Image-based mobile mapping systems enable an efficient acquisition of georeferenced image sequences, which can be used for geo-data capture in subsequent steps. In order to provide accurate measurements in a given reference frame while e.g. aiming at high fidelity 3D urban models, high quality georeferencing of the captured multi-view image sequences is required. Moreover, sub-pixel accurate orientations of these highly redundant image sequences are needed in order to optimally perform steps like dense multi-image matching as a prerequisite for 3D point cloud and mesh generation. While direct georeferencing of image-based mobile mapping data performs well in open areas, poor GNSS coverage in urban canyons aggravates fulfilling these high accuracy requirements, even with high-grade inertial navigation equipment. Hence, we conducted comprehensive investigations aiming at assessing the quality of directly georeferenced sensor orientations as well as the expected improvement by image-based georeferencing in a challenging urban environment. Our study repeatedly delivered mean trajectory deviations of up to 80 cm. By performing image-based georeferencing using bundle adjustment for a limited set of cameras and a limited number of ground control points, mean check point residuals could be lowered from approx. 40 cm to 4 cm. Furthermore, we showed that largely automated image-based georeferencing is capable of detecting and compensating discontinuities in directly georeferenced trajectories.

### 1. INTRODUCTION

In recent years, image-based mobile mapping has evolved into a highly efficient and accurate mapping technology as it enables capturing an enormous amount of metric image data in a short time period with no or just minimal road traffic interference. In contrast to airborne nadir applications, where ground sampling distance (GSD) remains constant over the complete mapping area, vehicle-based mobile mapping imagery shows large variations due to different distances to mapping objects. Hence, in our scenario, one pixel corresponds to 2-6 mm in object space for a typical measurement range of 4-14 m and it is 1 cm at 23 m. While infrastructure management applications often demand 3D measurement accuracies better than 10 cm, urban modelling requires absolute accuracies at the cm level. As earlier studies show, these requirements could already be met in open areas. Burkhard et al. (2012) obtained absolute 3D point measurement accuracies of 4-5 cm in average to good GNSS conditions using their stereovision mobile mapping system. The feasibility of the StreetMapper LiDAR mobile mapping system to produce dense 3D measurements at an accuracy level of 3 cm in good GNSS conditions was demonstrated by Haala et al. (2008). A different version of this system is described in Puente et al. (2013) who present an analysis on the current performance of several mobile terrestrial laser scanning systems. Nonetheless, GNSS conditions of land-based mobile mapping vehicles are often deteriorated by multipath effects and by shading of the signals caused by trees and buildings which aggravate fulfilling the accuracy requirements by direct georeferencing. Furthermore, distances between cameras and measured objects are typically a few meters, compared to several hundred meters for airborne applications. Therefore, the contribution of the GNSS positioning error to the overall error

budget is much larger than the contribution of the error from the attitude determination. Since airborne surveys are much less affected by GNSS degradations experienced by ground-based mobile mapping systems, Nebiker et al. (2012) proposed the fusion of ground-based imagery from mobile mapping systems with aerial imagery. First experiments showed horizontal accuracies in the order of 5 cm, equivalent to the ground sampling distance of the aerial imagery, and vertical accuracies of approx. 10 cm.

One of the main features of our mobile mapping system is the application of multiple cameras, which are used for dense 3D data capture applying the multi-view-stereo matching described in Cavegn et al. (2015). Such a configuration especially requires a high quality relative orientation of the image sequences. Preferably this is available to a sub-pixel level in order to efficiently apply coplanarity constraints during dense stereo matching. Furthermore, exploiting accurately co-registered highly redundant multi-view image sequences can lead to an improvement in accuracy, reliability as well as completeness of the resulting products such as depth maps, 3D point clouds and meshes.

Integrated georeferencing is inevitable in built-up urban environments with extended areas of poor GNSS coverage, especially when data captured at different days and daytimes – which is typical for city-wide mapping – needs to be combined (Nebiker et al., 2015). In contrast to direct georeferencing, which exclusively relies on the position and attitude information provided by the GNSS/INS system, integrated georeferencing often uses pre-computed camera positions and additionally exploits image observations in a bundle adjustment. Ellum & El-Sheimy (2006) proposed to feed coordinate updates (CUPTs) determined by photogrammetric bundle adjustment back into a

loosely coupled GNSS Kalman filter. This approach incorporating additional stereovision-based position updates was later on exploited by Eugster et al. (2012). Whereas they demonstrated a consistent improvement of the absolute 3D measurement accuracy from several decimeters to a level of 5–10 cm for land-based mobile mapping, Ellum & El-Sheimy (2006) achieved no improvement in mapping accuracy. Bayoud (2006) developed a SLAM system which does not rely on GNSS observations, but is solely based on inertial observations and tie points from vision sensors. Vision updates for position and orientation are used as external measurements in an inertial Kalman filter. The filtered positions and orientations are subsequently employed in the photogrammetric intersection to map the surrounding features which are used as control points for the resection in the next epoch. Hassan et al. (2006) perform bundle adjustment incorporating camera positions and orientations provided by a Kalman filter. In poor GNSS areas, weights of camera positions and orientations are small and hence the solution will only depend on image observations, which results in photogrammetric bridging. Similar approaches were also developed by Forlani et al. (2005) and Silva et al. (2014) in order to bridge land-based mobile mapping stereo image sequences in GNSS denied areas.

In the following, we first present our mobile mapping platform and test scenario in section 2. Section 3 briefly describes the calibration process followed by direct and image-based georeferencing. A systematic study aiming at assessing the quality of directly georeferenced sensor orientations in a challenging urban environment with frequent GNSS degradations is presented in section 4 and section 5 gives further results on the potential and quality of image-based georeferencing.

## 2. MOBILE MAPPING PLATFORM AND TEST SCENARIO

In order to enable georeferencing accuracy investigations, two mobile mapping campaigns incorporating different sensor constellations were carried out. These campaigns including sensor specifications as well as our test site are introduced in the following two sections.

### 2.1 Mobile mapping system

All data used for the investigations presented in this paper was captured by the multi-sensor stereovision mobile mapping system of the Institute of Geomatics Engineering (IVGI), University of Applied Sciences and Arts Northwestern Switzerland (FHNW). Although two campaigns with different sensor constellations were performed, the following investigations focus on data from the main stereovision system, which consists of two 11 MP cameras and a calibrated stereo base of 905 mm. These stereo cameras have a resolution of 4008 x 2672 pixels at a pixel size of 9  $\mu\text{m}$ , a focal length of 21 mm and resulting fields-of-view of 81° in horizontal and 60° in vertical direction. The multi-camera configuration is completed by additional HD cameras with a resolution of 1920 x 1080 pixels, a pixel size of 7.4  $\mu\text{m}$ , a focal length of 8 mm and a field of view of 83° x 53°. While two additional stereo bases with this type of cameras were established for the campaign in July 2014 (Cavegn et al., 2015), just a third HD camera was setup in the middle of the 11 MP cameras for the campaign in August 2015 (see Figure 1).

To enable direct georeferencing of the imagery acquired at typically 5 fps, a NovAtel SPAN inertial navigation system is used. The navigation system consists of a tactical grade inertial measurement unit featuring fiber-optics gyros of the type UIMU-LCI and a L1/L2 GNSS kinematic antenna. In case of good GNSS coverage, these sensors provide an accuracy of horizontally 10 mm and vertically 15 mm during post-processing (NovAtel, 2016). Accuracies of the attitude angles roll and pitch are specified with 0.005° and heading with 0.008°. A GNSS outage of 60 seconds lowers the horizontal accuracy to 110 mm and the vertical to 30 mm.



Figure 1. Sensor configuration of the IVGI mobile mapping system for the campaign in August 2015

### 2.2 Test area and test data

The test site depicted in Figure 2 is located at a very busy junction of five roads in the city center of Basel, Switzerland. It includes large and rather tall commercial properties which create a very challenging environment for GNSS positioning. Three street sections of this test site were mapped three times, once in July 2014 and twice during a day in August 2015, which is a difference in time of 13 months (see Table 1). In all nine cases data acquisition was performed shortly before noon and at good weather conditions. For our investigations we used 85 up to 191 stereo image pairs from the forward facing stereovision system on a sequence length between 108 m and 217 m. An along-track distance between successive image exposures of 1 m was targeted, but larger distances occurred at velocities higher than 18 km/h since the maximum frame rate was 5 fps.

Whereas the campaign in July 2014 was part of a complete survey of the city-state of Basel, the campaign in August 2015 was specifically performed for the investigations at our test site (see Figure 3). In order to capture optimal trajectories, we acquired kinematic data according to best practice as specified by the manufacturer. First, static initialization for approx. 3 minutes in an open sky area followed by levelling until approaching the test site was carried out. After the first mapping of the test site, an additional loop was driven so that data could again be acquired in the test site area. Returning to the start area, imagery was captured on our outdoor calibration field for the purpose of boresight alignment (Burkhard et al., 2012). A further loop served for levelling and there was a static observation at the end of around 4 minutes nearby the FHNW building as well. The GNSS station on its roof which is part of the Automated GNSS Network for Switzerland (AGNES) was defined as base station. The complete campaign resulted in a total trajectory length of 22.756 km and 12220 stereo image pairs acquired on 20.8.2015 from 10:17:53 until 11:19:29.

For our investigations we determined 51 points mainly on corners of road markings at an absolute 3D accuracy better than 1cm by tachymetry. They served either as ground control (GCP) or check points (CP) (see Figure 2 and Figure 17).

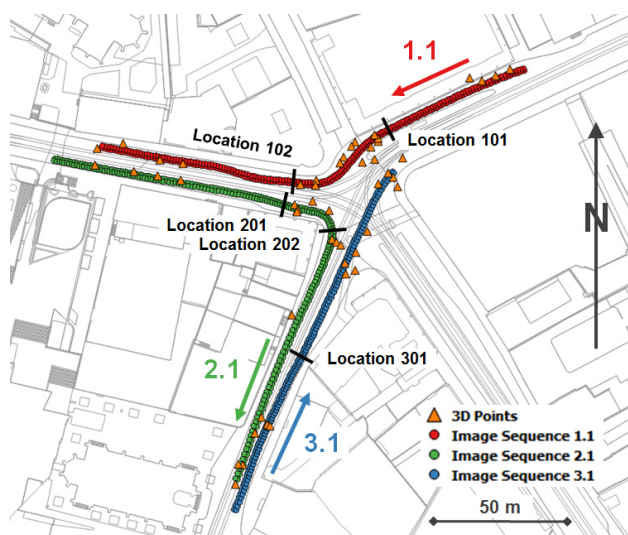


Figure 2. Base map of the test area with overlaid projection centers of selected stereo image sequences, 3D reference points and locations of trajectory discontinuities (Source: Geodaten Kanton Basel-Stadt)

Seq.	Date and Time	Image count	Length [m]	Along-track spacing [m]	
				Mean	Max.
1.0	24.7.14 10:20	246	164	1.34	1.97
1.1	20.8.15 10:30	322	173	1.08	1.25
1.2	20.8.15 10:47	312	175	1.13	1.48
2.0	27.7.14 11:53	314	173	1.11	1.60
2.1	20.8.15 10:34	342	212	1.25	2.06
2.2	20.8.15 10:50	382	217	1.14	1.93
3.0	27.7.14 11:57	170	108	1.29	1.49
3.1	20.8.15 10:37	232	141	1.23	1.73
3.2	20.8.15 10:53	192	146	1.54	2.37

Table 1. Characteristics of the nine selected stereo image sequences x.y (where x corresponds to the street sections 1 to 3 shown in Figure 2 and y corresponds to the campaign, 0 = 24./27.7.2014, 1 = 20.8.2015 10:30-10:37, 2 = 20.8.2015 10:47-10:53)

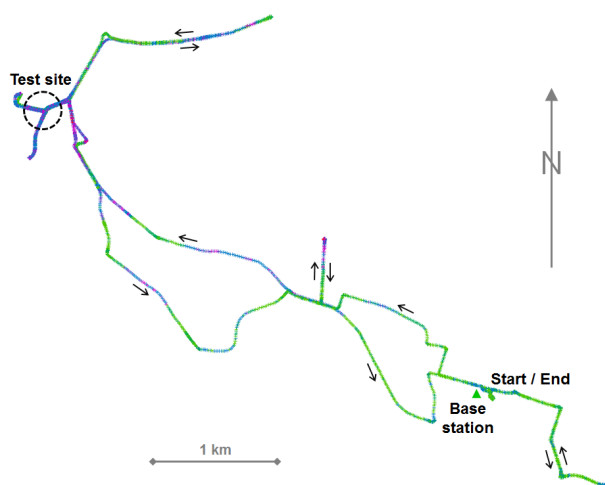


Figure 3. Trajectory of campaign on 20.8.2015 (green: high quality, red: low quality, test site: medium to low quality, trajectory extent in east-west direction is around 4.750 km)

### 3. CALIBRATION AND TRAJECTORY PROCESSING

A brief description of the calibration process followed by direct and image-based georeferencing is given in the following sections. Further investigations on the accuracy of the trajectories are then presented in chapter 4.

#### 3.1 System calibration

All sensors mounted on the rigid frame of the mobile mapping system were calibrated in an extensive and rigorous process. First, interior as well as relative orientation parameters between all cameras were determined by constrained bundle adjustment exploiting imagery taken on different indoor calibration fields for the two campaigns. While the indoor calibration field for the campaign in July 2014 features a uniform 3D point distribution, the indoor calibration field for the campaign in August 2015 does not have any 3D points on the ground. However, in both cases, many 3D points are signalized with coded targets. Second, lever arm and misalignment to the left camera of the forward looking stereo system were computed using forward imagery which was captured on our outdoor calibration field (Burkhard et al., 2012).

#### 3.2 Direct georeferencing

Navigation data was processed in tightly coupled mode using the GNSS and inertial post-processing software Inertial Explorer (version 8.60.4609) from NovAtel. Furthermore, processing was performed in multi-pass directions and trajectories were additionally smoothed. The resulting trajectory quality is depicted in Figure 3 and Table 2 shows that the estimated maximum 3D position accuracy values for all nine sequences are lower than 18 cm. By incorporating the previously computed boresight alignment as well as the relative orientation parameters, directly georeferenced sensor orientations were calculated for all images.

#### 3.3 Image-based georeferencing

Image-based georeferencing by bundle adjustment for each of the nine stereo image sequences was performed using Agisoft PhotoScan (version 1.2.3). Exterior orientation parameters from direct georeferencing as well as automatically determined image observations to tie points and manually defined image observations to approx. 20 ground control points per sequence were incorporated in the bundle adjustment. Even though input imagery was previously corrected for distortion and principal point based on the calibration parameters, significant radial distortion parameters were still estimated by bundle adjustment for the six sequences captured in August 2015 and were hence considered. The suboptimal point distribution in the calibration imagery could be a reason for the remaining distortion residuals of up to approx. 10 pixels in the image corners. This is depicted by Figure 4 where all eight calibration images per camera containing the exploited points and their corresponding residuals are overlaid.

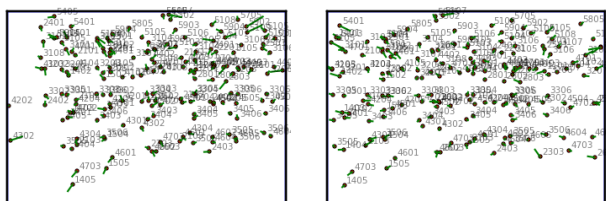


Figure 4. Point distribution of left and right stereo calibration imagery for the campaign in August 2015

#### 4. INVESTIGATIONS OF TRAJECTORY ACCURACY

The following two sections present a systematic study aiming at assessing the quality of directly georeferenced sensor orientations as well as its improvement by image-based georeferencing in a challenging urban environment with frequent GNSS degradations.

##### 4.1 Trajectory and orientation deviations between direct and image-based georeferencing

Deviations of projection centers and orientation angles between direct and integrated georeferencing were computed for all nine sequences (see Table 2 and Table 3). These 3D deviations range from 46 to 803 mm and the height is the component with the largest residuals for all sequences but for 3.2. Rather small deviations were obtained for street section 3. Sequence 1.1 shows the largest orientation deviations with ca. 0.5° for all orientation angles. However, a mean omega-phi-kappa deviation value of 0.3° was achieved.

Table 2 further depicts the maximum 3D position standard deviation per sequence estimated by GNSS/INS processing. The 3D mean value of deviations between direct and image-based georeferencing amounts to approx. 40 cm which is 3-4 times larger than the estimation by GNSS/INS based direct georeferencing. Moreover, sequence 1.2 with an improvement factor of more than 7 shows the largest discrepancy. Furthermore, it is clearly visible in Figure 5 that for all except of two sequences the RMSE values are larger than the maximum 3D position estimates by direct georeferencing. To sum up, the consistently overoptimistic 3D position accuracy estimates for direct georeferencing as provided from the NovAtel software proved not to be reliable for our data captured in urban environments.

Seq.	Direct (DSO) - Image-based (ISO)				Direct (DSO)	
	$\Delta E$ RMSE [mm]	$\Delta N$ RMSE [mm]	$\Delta h$ RMSE [mm]	$\Delta 3D$ RMSE [mm]	Max. 3D SD [mm]	Impr. fact.
1.0	298	36	423	519	140	3.7
1.1	41	86	126	158	138	1.1
1.2	437	26	572	721	100	7.2
2.0	52	37	75	98	65	1.5
2.1	264	83	498	569	86	6.6
2.2	172	470	628	803	128	6.3
3.0	23	30	77	86	172	0.5
3.1	15	25	36	46	52	0.9
3.2	172	500	146	548	132	4.2
Mean	164	144	287	394	113	3.6

Table 2. RMSE values for deviations of projection centers between direct and image-based georeferencing (RMSE DSO-ISO) as well as maximum 3D position estimates from direct georeferencing (Max. 3D SD DSO)

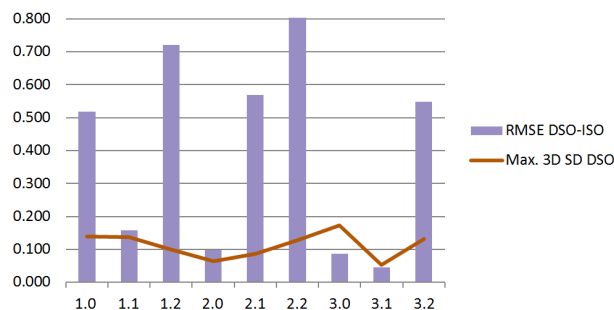


Figure 5. Comparison of RMSE values in m for 3D deviations of projection centers between direct and image-based georeferencing as well as maximum 3D position estimates from direct georeferencing

Seq.	$\Delta\omega$ RMSE [°]	$\Delta\phi$ RMSE [°]	$\Delta\kappa$ RMSE [°]	$\Delta\omega\phi\kappa$ RMSE [°]
1.0	0.246	0.095	0.246	0.361
1.1	0.366	0.072	0.354	0.515
1.2	0.311	0.067	0.295	0.434
2.0	0.143	0.056	0.160	0.222
2.1	0.319	0.081	0.298	0.445
2.2	0.310	0.064	0.283	0.425
3.0	0.011	0.055	0.008	0.057
3.1	0.088	0.079	0.047	0.127
3.2	0.084	0.057	0.048	0.112
Mean	0.209	0.070	0.193	0.300

Table 3. Deviations of orientation angles between direct and image-based georeferencing

Potential improvements in deviations of projection centers and hence in trajectory accuracy from integrated and image-based georeferencing over direct georeferencing is demonstrated in detail by Figures 6-14. Trajectories of stereo image sequences captured on the same street section at different times show differences of up to several decimeters. While small deviations were obtained for sequences 1.1 and 2.0, they are significantly larger for the other sequences of these two street sections. All deviations of street section 3 are smaller than 10 cm, with the exception of the north component of sequence 3.2 which amounts to approx. 50 cm. The order of the components is the same for each sequence of street section 1, i.e. east-north-height from positive to negative deviations. However, for the other street sections, no clear trend is visible.

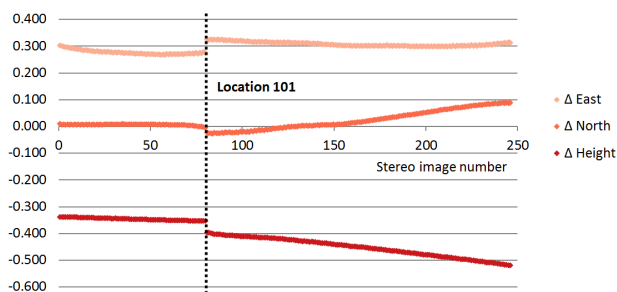


Figure 6. Deviations of projection centers in m for sequence 1.0

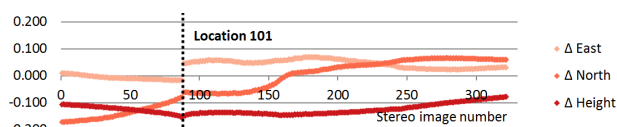


Figure 7. Deviations of projection centers in m for sequence 1.1



Figure 8. Deviations of projection centers in m for sequence 1.2

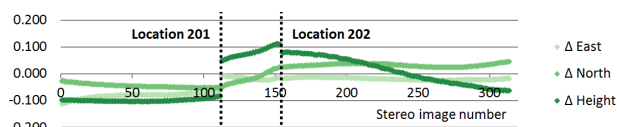


Figure 9. Deviations of projection centers in m for sequence 2.0



Figure 10. Deviations of projection centers for sequence 2.1

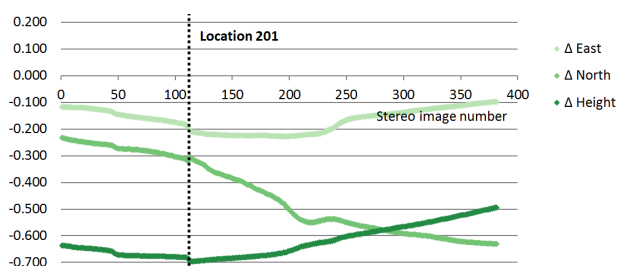


Figure 11. Deviations of projection centers for sequence 2.2

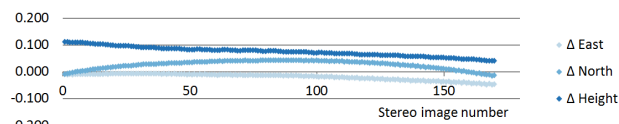


Figure 12. Deviations of projection centers in m for sequence 3.0

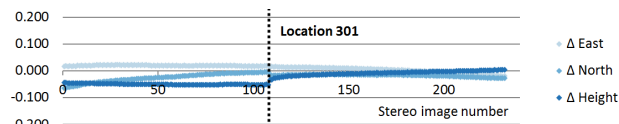


Figure 13. Deviations of projection centers in m for sequence 3.1

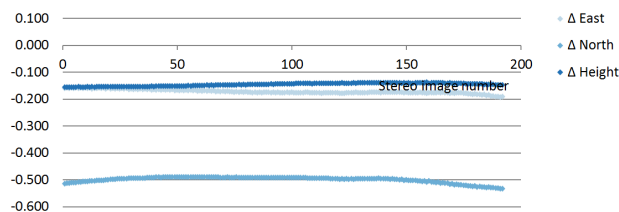


Figure 14. Deviations of projection centers in m for sequence 3.2

#### 4.2 Trajectory discontinuities from direct georeferencing

The charts illustrating trajectory deviations between direct and image-based georeferencing reveal nine trajectory discontinuities which are indicated by vertical dotted lines (see Figures 6-14). According to Figure 2 and Figure 15 all of them but one (location 202) were caused by a vehicle stop of several seconds mainly in front of crosswalks. However, no correlation between stop duration and 3D value of the discontinuities could be proven (see Table 4). 3D discontinuities amount mostly to a few centimeters, but they reach up to approx. 15 cm for sequence 2.0 at location 201. No discontinuities are present in sequences 3.0 and 3.2 since there were no stops. Furthermore, tuning the automated ZUPT detection tolerances in the GNSS/INS post-processing software Inertial Explorer might eliminate the trajectory discontinuities, but not the observed large systematic offsets.

Loc.	Seq.	Stop [s]	$\Delta E$ [mm]	$\Delta N$ [mm]	$\Delta h$ [mm]	$\Delta 3D$ [mm]
101	1.0	17	48	-21	-43	68
101	1.1	13	62	19	4	65
101	1.2	31	17	-27	-76	82
102	1.2	68	-3	2	-29	29
201	2.0	19	57	4	132	144
201	2.1	56	17	-3	46	49
201	2.2	41	-16	16	-12	26
202	2.0	0	-5	0	-30	30
301	3.1	16	-3	-12	18	22

Table 4. Dimensions of trajectory discontinuities

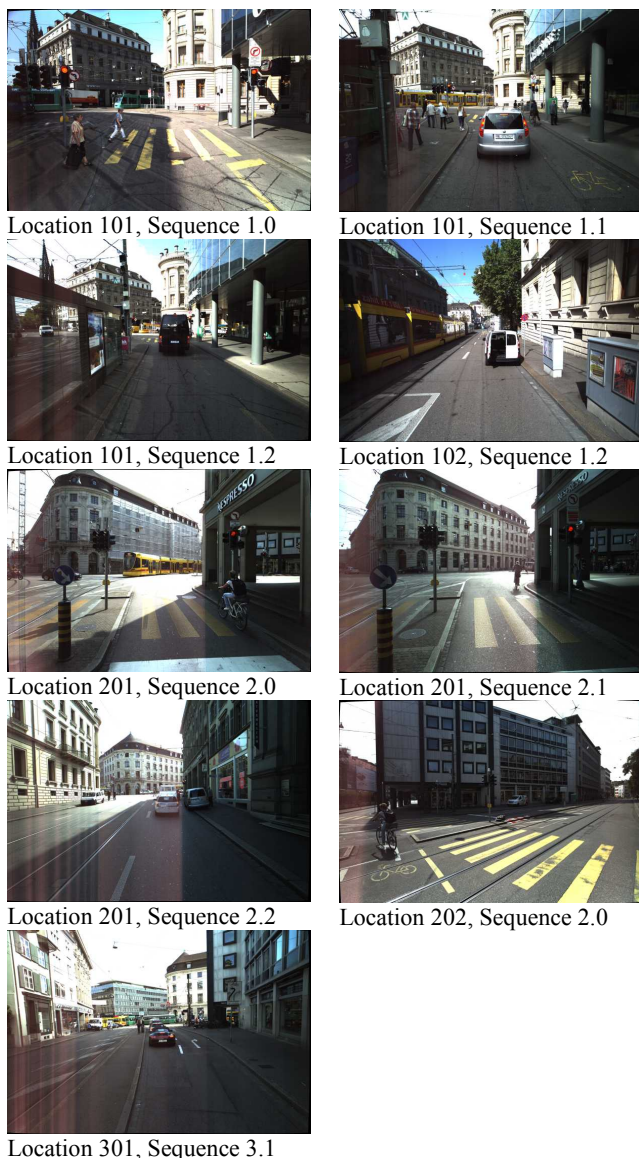


Figure 15. Mobile mapping imagery captured at locations of trajectory discontinuities

## 5. ACCURACY OF IMAGE-BASED GEOREFERENCING

In the following sections, accuracy of tie points and ground control points provided by bundle adjustment is discussed. Moreover, check point accuracy investigations for both direct and image-based georeferencing are presented.

### 5.1 Evaluation of bundle adjustment results

As described in section 3.3, bundle adjustment was performed by Agisoft PhotoScan using exterior orientation parameters from direct georeferencing as well as several ground control points. Since we newly set tie point accuracy to 0.3 pixel and defined 0.5 pixel for image observations to ground control points, also sequences acquired in July 2014 were reprocessed which led to slightly different results compared to Cavegn et al. (2015) and to Nebiker et al. (2015). Overall RMSE values of 0.42-0.89 pixel were computed, 0.15-0.21 pixel for tie points and 0.81-1.08 pixel for ground control points (see Table 5). The

resulting mean reprojection error for GCP of approx. 1 pixel is plausible if considering the samples depicted in Figure 16. Potential problems in 3D accuracy could be caused by the rather challenging identification and measurement of these natural ground control points e.g. compared to signalized targets. Other issues are varying distances to the 3D points mainly on corners of road markings leading to different object resolutions, e.g. for white strips or crosswalks. Whereas most of the residuals for the GCP of stereo image sequence 2.1 are smaller than 2 cm, the highest value amounts to 19 cm i.e. 1.9 pixel (see second sample on top row of Figure 16) which partly contributes to the largest 3D RMSE value of 47 mm (see Table 5). The mean tie points reprojection error of 0.18 pixel stands for relative orientations of high quality. Thus, for our scenario, the mean RMSE value of manually measured pixel coordinates at ground control points is larger by a factor of five if compared to automatic tie point measurements, while standard applications frequently assume a factor of two.

Seq.	Overall RMSE [px]	Tie points RMSE [px]	GCP RMSE [px]	GCP 3D RMSE [mm]
1.0	0.50	0.15	0.93	27
1.1	0.64	0.19	0.86	17
1.2	0.62	0.20	0.91	17
2.0	0.42	0.15	0.81	31
2.1	0.75	0.19	0.90	47
2.2	0.81	0.21	0.93	26
3.0	0.44	0.15	1.08	35
3.1	0.78	0.20	1.03	18
3.2	0.89	0.21	0.83	17
Mean	0.65	0.18	0.92	26

Table 5. Reprojection errors and 3D residuals of ground control points (GCP) from bundle adjustment

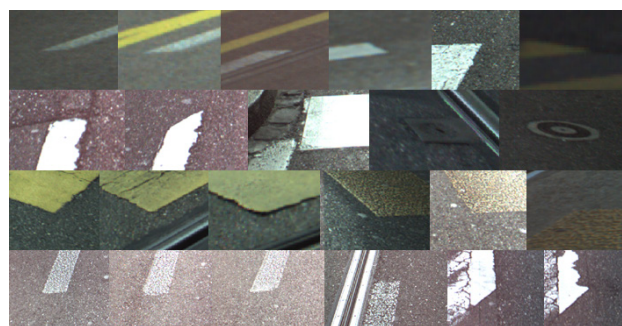


Figure 16. A sample for all 3D points of stereo image sequence 2.1 showing difficult identification

### 5.2 Check point investigations for direct and image-based georeferencing

In chapter 4.1 the 3D coordinates of camera trajectories both from direct georeferencing and bundle block adjustment were compared. While this gives some hints on the respective quality, accuracy investigations on 3D coordinates of measured image points are much more evident. For computation of these 3D coordinates by spatial intersection, orientation parameters both from direct georeferencing and bundle block adjustment can be used. Therefore, several groups of two, three or four ground control points (GCP) were established and approx. half of the previously used GCP were defined as check points (see Figure 17). Then, check point (CP) residuals were computed for two scenarios. First, only one GCP group at each end of a segment

was defined. Second, two additional GCP groups in-between and close to the corresponding sharp curve were established. For scenario one with two GCP groups, mean check point residuals of around 15 cm were obtained which is roughly 3 times better than a value of approx. 40 cm for direct georeferencing (see Table 6). Scenario two featuring four GCP groups, which led to check point residuals per sequence of 21-73 mm, shows to improve the direct georeferencing accuracy by an order of magnitude.

Seq.	CP count	Direct	Image-based 2 GCP groups		Image-based 4 GCP groups	
		$\Delta 3D$ [mm]	$\Delta 3D$ [mm]	Impr. fact.	$\Delta 3D$ [mm]	Impr. fact.
1.0	15	555	137	4.1	27	20.4
1.1	11	168	42	4.0	21	8.0
1.2	11	774	121	6.4	26	29.4
2.0	11	131	76	1.7	48	2.7
2.1	12	593	432	1.4	73	8.1
2.2	11	813	425	1.9	36	22.5
3.0	8	174	42	4.2		
3.1	10	64	30	2.1		
3.2	10	568	53	10.8		
Mean		427	151	2.8	39	11.1

Table 6. RMSE values for check point residuals of direct and image-based georeferencing

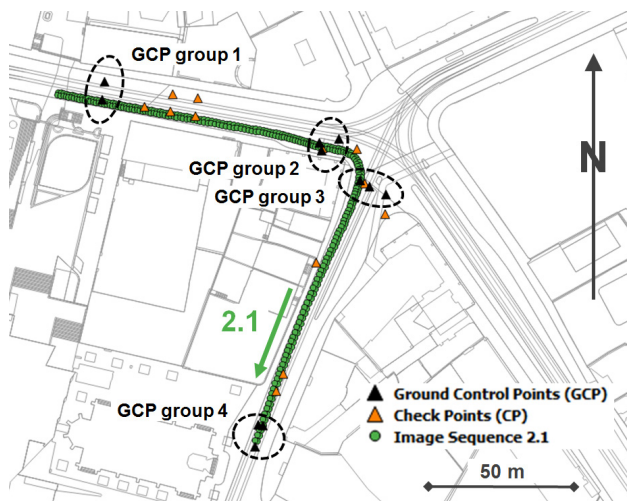


Figure 17. Locations of ground control point groups as well as check points for stereo image sequence 2.1 (Source: Geodaten Kanton Basel-Stadt)

## 6. CONCLUSIONS AND OUTLOOK

Comprehensive investigations exploiting high quality reference and temporal image data in our urban test site repeatedly led to 3D trajectory accuracies from direct georeferencing in the order of one to several decimeters. By performing image-based georeferencing using bundle adjustment of only the forward stereo imagery and a number of ground control points, object point accuracies of approx. 4 cm were consistently obtained which is an improvement by an order of magnitude. Since image observations over the entire sequences were considered, not only large offset and drift errors from direct georeferencing were removed, but also trajectory discontinuities could be detected and compensated.

Achieved absolute 3D point accuracies in the order of a few centimeters are sufficient for typical infrastructure management applications. However, 3D measurements were only based on forward stereo imagery and a considerable number of ground control points is still required. Efficiency and accuracy will be increased by incorporating multi-view stereo image sequences into bundle adjustment exploiting constraints for the calibrated offsets and rotations between respective cameras. By enabling self-calibration, which has been a standard procedure in the airborne case for many years, even remaining inaccuracies from suboptimal test field calibration could be compensated. Our study demonstrated manual image measurement accuracies of 1 pixel for objects of interest which fulfills urban mapping requirements. Moreover, the achieved sub-pixel accurate relative orientations are sufficient in order to perform dense multi-image matching. Nonetheless, due to completely different views, automated tie point detection and matching in mobile mapping sequences pose one of the biggest challenges which will be met in future research. Furthermore, aiming at processing complete cities, a new integrated and image-based georeferencing approach will be developed which can handle multiple large image sequences.

## 7. ACKNOWLEDGEMENTS

Thanks are due to iNovitas AG (Baden-Dättwil, Switzerland) and to the City of Basel (Bau- und Verkehrsdepartement Kanton Basel-Stadt) for providing some of the mobile mapping data. This work was co-financed by the Swiss Commission for Technology and Information CTI as part of the infraVIS project.

## 8. REFERENCES

- Bayoud, F. A., 2006. Development of a Robotic Mobile Mapping System by Vision-Aided Inertial Navigation: A Geomatics Approach. PhD thesis, EPF Lausanne.
- Burkhard, J., Cavegn, S., Barmettler, A. & Nebiker, S., 2012. Stereovision Mobile Mapping: System Design and Performance Evaluation. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B5, pp. 453-458.
- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. & Zwölfer, T., 2015. Evaluation of Matching Strategies for Image-Based Mobile Mapping. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, La Grande Motte, France, Vol. II-3/W5, pp. 361-368.
- Ellum, C. M. & El-Sheimy, N., 2006. New Strategies for Integrating Photogrammetric and GNSS Data. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Dresden, Germany, Vol. XXXVI, Part 5.
- Eugster, H., Huber, F., Nebiker, S. & Gisi, A., 2012. Integrated Georeferencing of Stereo Image Sequences Captured with a Stereovision Mobile Mapping System – Approaches and Practical Results. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B1, pp. 309-314.

Forlani, G., Roncella, R. & Remondino, F., 2005. Structure and Motion Reconstruction of Short Mobile Mapping Image Sequences. In: *VII Conference on Optical 3D Measurement Techniques*, Vienna, Austria, pp. 265-274.

Haala, N., Peter, M., Kremer, J. & Hunter, G., 2008. Mobile LiDAR Mapping for 3D Point Cloud Collection in Urban Areas - A Performance Test. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Beijing, China, Vol. XXXVII, Part B5, pp. 1119-1124.

Hassan, T., Ellum, C. & El-Sheimy, N., 2006. Bridging Land-Based Mobile Mapping Using Photogrammetric Adjustments. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Paris, France, Vol. XXXVI, Part 1.

Nebiker, S., Cavegn, S., Eugster, H., Laemmer, K., Markram, J. & Wagner, R., 2012. Fusion of Airborne and Terrestrial Image-Based 3D Modelling for Road Infrastructure Management – Vision and First Experiments. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B4, pp. 79-84.

Nebiker, S., Cavegn, S. & Loesch, B., 2015. Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *ISPRS International Journal of Geo-Information*, 4(4), pp. 2267-2291.

NovAtel, 2016. SPAN UIMU-LCI. <http://www.novatel.com/assets/Documents/Papers/IMU-LCI.pdf> (6.4.2016).

Puente, I., González-Jorge, H., Martínez-Sánchez, J. & Arias, P., 2013. Review of Mobile Mapping and Surveying Technologies. *Measurement*, 46(7), pp. 2127-2145.

Silva, J. F. C., Lemes Neto, M. C. & Blasechi, V., 2014. Automating the Photogrammetric Bridging Based on MMS Image Sequence Processing. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Denver, USA, Vol. XL-1, pp. 389-396.